Identification of errors-in-variables model with observation outliers based on MCD

J. ALMutawa
Department of Applied mathematics and physics
Kyoto University, Kyoto, 606-8501, Japan
e-mail:{jaafar}@amp.i.kyoto-u.ac.jp

***Abstract*— In this paper, we develop a subspace system identification algorithm for the Errors-In-Variables (EIV) model subject to observation noise with outliers. To this end, we proposed the random search algorithm in order to solve the Minimum-Covariance-Determinant (MCD) problem. By using the MCD, we identify and delete the outliers, and then we apply the classical EIV subspace system identification algorithms to get state space model. In addition, we show that the problem of detecting the outliers in the closed loop systems is especial case of the EIV model. The propose algorithm has been applied to heat exchanger data.**

Keywords: Subspace system identification, errors-in-variables model, outliers, Minimum-Covariance-Determinant, random search algorithm.

I. Introduction

A basic numerical procedure in the EIV or closed-loop subspace system identification methods e.g. MOESP, [3], N4SID, [4], is based on the LQ-decomposition, which essentially computes the least-squares estimate of the future based on the past. Since the least squares method is rather sensitive to outliers (non Gaussian disturbances) so are the subspace system identification methods. Moreover, it is well known in real applications that there are some cases where large errors are contained in observation data with low probability, so that a standard Gaussian assumption for observation noises may fail.

In view of this observation, we have already presented a subspace system identification method subject to observation outliers [5], in which a subspace system identification method coupled with a weighted least-squares (WLS) method is derived to iteratively compute the state space model. This method assumes that the outliers occurs only in the output data, however, usually the input data are observed quantities subject to random variability. Therefore there is no reason why gross errors would only occur in the response data. In a certain sense it is more likely to have outlier in one of observed input data, because usually its dimension greater than one, and hence there are more opportunities for some thing to go wrong. For the case where the outliers acts in the closed loop system or occur in the input data, to the best of out knowledge there is no paper has been published in the area of subspace system identification methods.

In this paper, we consider a subspace system identification problem in the presence of observation outliers with the aid of the MCD procedure. Two cases has been studied, EIV and closed loop systems with outliers. To this end, we proposed the random search algorithm [1] to solve the MCD problem. Then, by using the MCD procedure, we derive a subspace system identification method which does not require an initial estimate of the parameters as in [5], [6]. It should be noted that the technique presented here can be viewed as a two stage procedure: Identification of outliers followed by the application of the subspace system identification method. The first stage can be considered as a preprocessing step.

Historically, the MCD criteria was first suggested by Rousseeum [8]. Then, [7] presented a fast MCD algorithm to compute MCD estimates for the multivariate linear regression model. We will propose a random algorithm to solve the MCD problem for the state space model; this algorithm has a unique feature in that if the input is a white noise and the number of data is very large, the algorithm becomes deterministic and provides the MCD estimate in a probabilistic sense. On the other hand, if the input is not white noise and the number of data is finite, then there is a confidence interval to compute the objective function, which can improve the estimate obtained by subspace algorithms.

This paper is organized as follows. Section 2 introduces the MCD problem for the multivariate linear regression model. In section 3, we state the errors-in-variables problem in the presence of outliers, and consequently the MCD problem has been solved using the random search algorithm. In section 4 we modify the MCD algorithm to detect the outliers for the closed loop. Section 5 considers the identification of noisy input-out data for the heat exchanger.

II. Minimum covariance determinant

Consider the multivariate linear regression model

$$y_i = \theta^T \phi_i + v_i, \quad i = 1, \cdots, \mathcal{N} \tag{1}$$

where $y_i \in \mathbb{R}^p$ is the output vector, $\phi_i \in \mathbb{R}^q$ the regressors vector, $\theta \in \mathbb{R}^{q \times p}$ contain the unknown regression coefficients, and $v_1, \cdots, v_{\mathcal{N}}$ are i.i.d. with zero mean and positive definite symmetric covariance matrix Σ of size p. The linear regression model (1) can be written in a matrix form as

$$\mathbf{Y} = \Phi\theta + \mathbf{v}, \tag{2}$$

where

$$\mathbf{Y} = \begin{bmatrix} y_1^T \\ \vdots \\ y_{\mathcal{N}}^T \end{bmatrix}, \; \Phi = \begin{bmatrix} \phi_1^T \\ \vdots \\ \phi_{\mathcal{N}}^T \end{bmatrix}, \; \mathbf{v} = \begin{bmatrix} v_1^T \\ \vdots \\ v_{\mathcal{N}}^T \end{bmatrix}. \tag{3}$$

The classical ordinary Least Squares (LS) estimator $\hat{\theta}_{LS}$ is defined by $\hat{\theta}_{\text{LS}} = (\Phi^T\Phi)^{-1}(\Phi^T\mathbf{Y})$, while the covraince Σ is unbiasedly estimated by $\hat{\Sigma}_{LS} = \frac{1}{\mathcal{N}-p}(\mathbf{Y} - \Phi\hat{\theta}_{LS})^T(\mathbf{Y} - \Phi\hat{\theta}_{LS})$ and the ith residual is defined as $r_i(\theta) = y_i - \hat{\theta}_{\text{LS}}^T\phi_i$.

In this paper we will study two type of outliers, the vertical and leverage outliers. A point (ϕ_i, y_i) is called vertical outliers if it does not follow the linear pattern of the majority of the data but whose ϕ_i is not outlying, while if ϕ_i is outlying w it is called leverage point. In general, it is difficult to identify the leverage points, because of the higher dimensionality. Moreover, it is not sufficient to look at each variable separately or even at all plots of pairs of variables. Many researchers will argue that regression outliers can be discovered by looking at the least squares of residuals. Unfortunately, this is not true. Consider for example Fig.

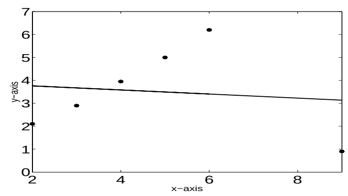

Fig. 1. Leverage point

1, the outlier residual has small negative number, on the other hand, some residuals have much larger absolute values, although they correspond to good points. Of course, in such a bivariate data set one can actually view the data, but there are many multivariate data sets where outliers remain invisible.

To this end we apply the MCD techinque to detect the outliers. To define the MCD, consider a data set $Z_{\mathcal{N}} = \{z_i = \begin{bmatrix} \phi_i \\ y_i \end{bmatrix} : i = 1, \cdots, \mathcal{N}\}$, and let $\mathcal{S} = \{S \subseteq \{1, \cdots, \mathcal{N}\} : \#S = M\}$ [1] be the collection of all subsets with cardinality equal M from the set $\{1, \cdots, \mathcal{N}\}$, where $[\mathcal{N}/2] \leq M \leq \mathcal{N}$. For any $S \in \mathcal{S}$, let $Z_S = \{z_i = \begin{bmatrix} \phi_i \\ y_i \end{bmatrix} : i \in S\}$, and define the location of the parameter as the empirical mean $\bar{\Phi}(S) = \frac{1}{M}\sum_{i\in S}^{M} \phi_i$, and scatter is the empirical covariance estimate $\text{cov}(S) = \frac{1}{M}\sum_{i\in S}^{M}(z_i - \bar{z}_i)(z_i - \bar{z}_i)^T$ where $\bar{z}_i = \begin{bmatrix} \bar{\Phi}(S) \\ \hat{y}_i \end{bmatrix}$ and $\hat{y}_i$ is the LS estimate based on Z_S. The MCD estimator is define as

$$\text{Minimized } \det(\text{cov}(S)) \quad \text{subject to}$$
$$\frac{1}{M}\sum_{i=1}^{M}\sqrt{(z_i - \bar{z}_i)^T\text{cov}(S)^{-1}(z_i - \bar{z}_i)} \tag{4}$$

[1] $\#$:= cardinality of the subset S.

i.e the MCD searches for the subset Z_S of size M whose covariance matrix has the smallest determinant subject to the squared Mahalanobis distance is $d_i^2 = (z_i - \bar{z}_i)^T\text{cov}(S)^{-1}(z_i - \bar{z}_i)$.

In the next section, we will formulate the MCD for the state-space model, and apply the random search algorithm to solve the MCD problem.

III. Errors-in-variables models

As depicted in Fig. 1, consider the errors-in-variables system described by

$$\begin{bmatrix} x_{t+1} \\ \tilde{y}_t \end{bmatrix} = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \begin{bmatrix} x_t \\ \tilde{u}_t \end{bmatrix}, \tag{5}$$

where $x_t \in \mathbb{R}^n$, $\tilde{u}_t \in \mathbb{R}^m$ and $\tilde{y}_t \in \mathbb{R}^p$. The measured

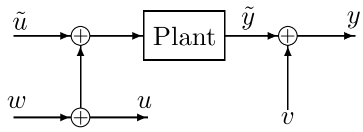

Fig. 2. Error-in-variables models

input-output signals y_t and u_t are modeled as

$$u_t = \tilde{u}_t + w_t, \tag{6}$$
$$y_t = \tilde{y}_t + v_t. \tag{7}$$

Assume a time sequence of data $\{(u_t, y_t),\ t = 1, \cdots, N + 2l - 1 = \mathcal{N}\}$ is given, the problem of interest is to estimate the quadruple (A, B, C, D) within the freedom of equivalent transformation. The system noise $w_t \in \mathbb{R}^n$ and the measurement noise $v_t \in \mathbb{R}^p$ are Non-Gaussian white noises that contain outliers with a low probability. Therefore we write:

$$w_t = (I_n - \text{diag}\{\gamma_{t,i}\})w_t^n + \text{diag}\{\gamma_{t,i}\}w_t^o,\ i = 1, \cdots, n, \tag{8}$$
$$v_t = (I_p - \text{diag}\{\alpha_{t,j}\})v_t^n + \text{diag}\{\alpha_{t,j}\}v_t^o,\ j = 1, \cdots, p, \tag{9}$$

where I_s is the $s \times s$ identity matrix for $s = \{n, p\}$, $\text{diag}\{\delta_{t,j}\} = \text{diag}\{\delta_{t,1}, \cdots, \delta_{t,p}\}$ and $\delta_{t,j} = \{0, 1\}$ where $\delta = \{\gamma, \alpha\}$. Moreover, $\text{Prob}\{\delta_{t,i} = 1\}$ is small, i.e. the majority of the observed data is clean. The noises $\{w_t^n, w_t^o, v_t^n,\ v_t^o\}$ are Gaussian white noises with

$$w_t^n \in \text{N}(0, Q^n), \qquad w_t^o \in \text{N}(\epsilon_1, Q^o), \tag{10}$$
$$v_t^n \in \text{N}(0, R^n), \qquad v_t^o \in \text{N}(\epsilon_2, R^o). \tag{11}$$

where $\{Q^n, Q^o, R^n, R^o\}$ are positive definite covariance matrices, where $Q_{i,i}^o$ and $R_{i,i}^o$ are much larger than $Q_{i,i}^n$ and $R_{i,i}^n$ respectively. In particular, ϵ_i for $i = 1, 2$ are vectors whose elements are small numbers that allows the outliers to change the mean. The fact that that we account for the possibility that the input signal is not exactly known and it may contain outliers, makes the problem difficult, and is often referred to as an outlier-errors-in-variables (OEIV) problem.

It should be noted that the outliers formulations (10-11), include mean shift outlier and variance-shift model. There are another plausible outliers formulations, as might be expected, alternative formulations can lead to different procedures.

Substituting (6) and (7) in (5) yields

$$\begin{bmatrix} x_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix} - \begin{bmatrix} Bw_t \\ Dw_t - v_t \end{bmatrix}$$

let $w_t^1 = -Bw_t$ and $v_t^1 = -Dw_t + v_t$

$$\operatorname{cov} \begin{bmatrix} w_t^1 \\ v_t^1 \end{bmatrix} = \begin{bmatrix} B & 0 \\ D & I \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} B & 0 \\ D & I \end{bmatrix}^T$$

and

$$\begin{bmatrix} x_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix} + \begin{bmatrix} w_t^1 \\ v_t^1 \end{bmatrix}$$

A. MCD for the error-in-variables models

Assume that $\hat{x}_t$ is given estimate of the state x_t for $t = 1, \cdots, \mathcal{N}$. Then the EIV model (5) is a linear regression model in the quadruple (A, B, C, D), where $\begin{bmatrix} \hat{x}_{t+1} \\ y_t \end{bmatrix}$ and $\begin{bmatrix} \hat{x}_t \\ u_t \end{bmatrix}$ are known input-output data of the regression model.

Proposition 1: Let $S \in \mathcal{S}$ and $z_t = \begin{bmatrix} \hat{x}_t \\ u_t \\ y_t \end{bmatrix}$, then the MCD cost function for the EIV model (5) can be written as

$$J(A, B, C, D) = \min_{A,B,C,D} \det \operatorname{cov}(S)$$
$$\text{subject to} \quad \frac{1}{M} \sum_{i=1}^{M} \sqrt{(z_t - T_t)^T \operatorname{cov}(S)^{-1} (z_t - T_t)} \qquad (12)$$

where $\operatorname{cov}(S) = \sum_{t \in S} \Big(z_t - T_t\Big)\Big(z_t - T_t\Big)^T$, and $T_t = \begin{bmatrix} \frac{1}{M} \sum_{t \in S} \hat{x}_t \\ \frac{1}{M} \sum_{t \in S} u_t \\ \hat{y}_t \end{bmatrix}$.

According to Proposition 1, $J(A, B, C, D)$ can be found by searching for the best subset $S \in \mathcal{S}$ that minimizes the $\det \operatorname{cov}(S)$. However there are $\binom{\mathcal{N}}{M}$ subsets in $\mathcal{S}$, so that finding the best subset that minimizes the value of the objective function is very difficult problem. In fact, for each subset $S \in \mathcal{S}$, we can easily evaluate the $\det \operatorname{cov}(S)$, then sorting $\det \operatorname{cov}(S)$ in increasing order, i.e.

$$\det \operatorname{cov}(S)_{[1]} = \min_{S \in \mathcal{S}} \det \operatorname{cov}(S) \leq \cdots \leq \det \operatorname{cov}(S)_{[\binom{\mathcal{N}}{M}]} = \max_{S \in \mathcal{S}} \det \operatorname{cov}(S).$$

Now we think of $S \in \mathcal{S}$ as a random variable who is uniformly distributed, and hence $\det \operatorname{cov}(S)$ is also a random variable depending on S. Let $F(\det \operatorname{cov}(S))$ denote the unknown probability distribution function of $\det \operatorname{cov}(S_i)$, for $i = 1, \cdots, L$ be L independently generated samples of $S \in \mathcal{S}$. Let $\bar{S} \in \{S_i\}_1^L$ such that $\det \operatorname{cov}(\bar{S}) = \min_{1 \leq i \leq L} \det \operatorname{cov}(S_i)$. The next theorem finds the sample size L so that $\det \operatorname{cov}(\bar{S})$ converge to the true solution with probability close to one.

Theorem 1: Consider the EIV model (5), then the following (i) $\sim$ (iii) hold:

(i) For all $0 < F(\min_{S \in \mathcal{S}} \det \operatorname{cov}(S)) < \epsilon < 1$ and for all $0 < \delta < 1$, if

$$L \geq \frac{\ln(1/\delta)}{\ln(1/(1-\epsilon))}, \qquad (13)$$

then

$$\text{Prob}\Big\{F\Big(\min_{1 \leq i \leq L} \det \operatorname{cov}(S_i)\Big) \leq \epsilon\Big\} \geq 1 - \delta.$$

(ii) $F(\min_{S \in \mathcal{S}} \det \operatorname{cov}(S))$ has a Beta distribution function [2].

(iii) The mean and the variance of $F(\min_{S \in \mathcal{S}} \det \operatorname{cov}(S, \theta))$ are respectively given by

$$\mu = \frac{1}{1 + \binom{\mathcal{N}}{M}}, \quad \sigma^2 = \frac{\binom{\mathcal{N}}{M}}{(1 + \binom{\mathcal{N}}{M})^2 (\binom{\mathcal{N}}{M} + 2)}.$$

Proof: (i) Let $\det \operatorname{cov}(S)_{[k]}$ denote the maximum $\det \operatorname{cov}(S)$ that satisfies $\mathrm{F}(\det \operatorname{cov}(S)) \leq \epsilon$, i.e

$$\mathrm{F}(\det \operatorname{cov}(S)_{[\mathcal{N}]}) \geq \cdots \geq \mathrm{F}(\det \operatorname{cov}(S)_{[k+1]}) > \epsilon$$

Now

$$\mathrm{F}\big(\min_{1 \leq i \leq L} \det \operatorname{cov}(S_i)\big) \leq \epsilon \Leftrightarrow \min_{1 \leq i \leq L} \det \operatorname{cov}(S_i) \leq \det \operatorname{cov}(S)_{[k]}$$

implies that

$$\begin{aligned} &\text{Prob}\Big\{\mathrm{F}\big(\min_{1 \leq i \leq L} \det \operatorname{cov}(S_i)\big) \leq \epsilon\Big\} \\ &= \text{Prob}\Big\{\min_{1 \leq i \leq L} \det \operatorname{cov}(S_i) \leq \det \operatorname{cov}(S)_{[k]}\Big\} \\ &= 1 - \text{Prob}\Big\{\min_{1 \leq i \leq L} \det \operatorname{cov}(S_i) > \det \operatorname{cov}(S)_{[k]}\Big\} \\ &= 1 - \text{Prob}\Big\{\det \operatorname{cov}(S_1) \geq \det \operatorname{cov}(S)_{[k+1]}\Big\} \cdots \\ &\quad \times \text{Prob}\Big\{\det \operatorname{cov}(S_L)) \geq \det \operatorname{cov}(S)_{[k+1]}\Big\} \\ &\geq 1 - (1-\epsilon)^L \end{aligned}$$

Now $\mathrm{L} \geq \frac{\ln(\frac{1}{\delta})}{\ln \frac{1}{(1-\epsilon)}} \Rightarrow (1-\epsilon)^L \leq \delta$.

$$\Rightarrow \text{Prob}\Big\{\mathrm{F}(\min \det \operatorname{cov}(S_i)) \leq \epsilon\Big\} \geq 1 - (1-\epsilon)^L \geq 1 - \delta$$

(ii) Let $\mathcal{J} = \{\det \operatorname{cov}(S)_{[i]} : i = 1, \cdots, \binom{\mathcal{N}}{M}\}$ be the set of all possible values of the objective function. Then $\det \operatorname{cov}(S)_{[i]}$ is uniformly distributed and hence the problem can be considered as order statistics. Therefore the order statistics will have a Beta distribution function this proves (ii). Hence (iii) is a trivial application of the Beta distribution function.

Theorem 1, means that, whenever we generate independent random subsets $S_L = \{S\}_{i=1}^L$ and computes the value of the objective function in Proposition 1 for each subset $S_i \in S_L$. Then, finding a subset $\bar{S} \in S_L$ with minimum covariance determinant under the constrain that this value of the objective function is smaller than or equal the LS estimate corresponding the whole data will improve our estimate.

[2] A variable x has a Beta distribution if and only if its probability density is given by

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

where Γ is the gamma function, $\alpha > 0$ and $\beta > 0$.

However, it may be noted that in the worst case this improvement is not considerable comparing to the LS estimate corresponding to the whole data. In fact, if the number of the observed data is very large, then the probability of finding a subset $S \in \mathcal{S}$ with cardinality equal to M that does not contain any outliers approaches zero. i.e.

$$P_I = \frac{\binom{I}{M}}{\binom{\mathcal{N}}{M}} = \frac{I!(\mathcal{N}-M)!}{(I-M)!\mathcal{N}!} = \prod_{j=0}^{M-1} \frac{I-j}{\mathcal{N}-j}, \qquad (14)$$

where $I = \epsilon\mathcal{N}$ stands for the number of clean data. According to (14) the random search algorithm can be improved by taking S with small cardinality then finding the smallest M relative Mahalanobis distances d_i. This will increase the probability of finding a subset S from $\mathcal{S}$ that does not contain any outliers. In the multivariate linear regression model (1), the subset S should be at least of cardinality equal to $p+1$, which is smaller than that we will not be able to estimate the unknown parameters. In the state space model, the cardinality $M > n$, where n is the dimension of the state x_t.

An advantage of Theorem 1, if we minimized the objective function in the norm form instead of the determinant. Then the random search will not have probability greater than $(1-\delta)$ succeeds, instead the probability of succeeds will be $(1-p\delta)$. So that for large p the random search will not be suitable, for example let $p = 10,\ \delta = 0.05$ so that the probability of succeeds will be greater than $1 - p\delta = 0.5$.

The disadvantage is that we can not measure the state x_t, so that Proposition 1 require computing the subspace system identification algorithms in order to search for the best subset that minimizes the MCD cost function, which will consume time. This can be avoided by preprocessing the observed output data, then after detecting and deleting the outliers, we compute the state space parameters via the well known subspace system identification algorithm e.g EIV-MOESP, or EIV-N4SID.

Theorem 2: Let $S = \{\pi_i : i = 1, \cdots, M\}$, then the MCD for the EIV model (5), can be written as

$$\min_{S\in\mathcal{S}} \det \operatorname{cov}(S) = \min_{S} \det \frac{1}{M} \sum_{i\in S} \left(\begin{bmatrix} u_i \\ y_i \end{bmatrix} - \begin{bmatrix} \frac{1}{M}\sum_{i\in S} u_i \\ \hat{y}_i \end{bmatrix} \right) \times \left(\begin{bmatrix} u_i \\ y_i \end{bmatrix} - \begin{bmatrix} \frac{1}{M}\sum_{i\in S} u_i \\ \hat{y}_i \end{bmatrix} \right)^T \qquad (15)$$

Proof: The EIV model (5) can be written as following

$$\tilde{Y}_S = x_{\pi_1}^T \mathcal{O} + \tilde{U}_S \Gamma, \qquad (16)$$

using (6) and (7) in (16) we get

$$Y_S = x_{\pi_1}^T \mathcal{O} + U_S\Gamma + V_S - W_S\Gamma, \qquad (17)$$

where

$$Y_S := \begin{bmatrix} y_{\pi_1}^T \\ \vdots \\ y_{\pi_M}^T \end{bmatrix}, \quad U_S^T := \begin{bmatrix} u_{\pi_1} \\ \vdots \\ u_{\pi_M} \end{bmatrix}, \quad \mathcal{O} = \begin{bmatrix} C^T \\ (CA)^T \\ \vdots \\ (CA^{\pi_M - 1})^T \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} D^T & (CB)^T & \cdots & (CA^{\pi_M-2}B)^T \\ 0 & D^T & \cdots & \vdots \\ \vdots & \ddots & & (CB)^T \\ 0 & \cdots & 0 & D \end{bmatrix}$$

Following the above theorem we avoid finding an initial estimate which is required in order to detect the outliers as in [5], [6]. Another advantage of the above algorithm is that we do not need to compute every time the subspace algorithms in order to find the best objective function.

Random search algorithm 1:

Choose the error bound ϵ, and the confidence interval δ,

- **Step 1:** Calculate the sample size L as in (13).
- **Step 2:** Randomly generate L samples of $S_i \in \mathcal{S},\ i = 1, \cdots, L$, according to uniform distribution. For each sample S_i, calculate $\det\operatorname{cov}(S_i)$ as in (15) and consequently find $\det\operatorname{cov}(S^+) = \min_{1\le i\le L} \det\operatorname{cov}(S_i)$.
- **Step 3:** Compute d_i for $i = 1, \cdots, \mathcal{N}$ and detect outliers using Chi-square distribution. Then, put $S_1 = \{\pi(1), \cdots, \pi(M)\}$.
- **Step 5:** Repeat step 1 to step 3, until convergent.
- **Step 6:** Formulate the block Hankel matrices by using the data in the set S_1 and compute the state space parameters via the subspace algorithms e.g. EIV-MOESP, EIV-N4SID, etc.

Although, when $\mathcal{N} \to \infty$, it seems that finding a subset $S \in \mathcal{S}$, that does not contain any outlier is hopeless, the next theorem finds a deterministic algorithm to approximate the value of the MCD objective function.

Theorem 3: Assume that $\gamma_{t,i} = 0$, for all t and i, and let u_i be white noise with zero mean, finite covariance matrix $\sigma^2 I$ and independent of u_j for every $i \neq j$. The value of the MCD is minimized whenever we arrange

$$\|y\|_{[1]}^2 \le \cdots \le \|y\|_{[M]}^2$$

where $\|y\|_{[i]}^2$ is the ith order statistics of $\|y_j\|$ for $j = 1, \cdots, \mathcal{N}$.

Proof: Ignoring the initial state in (16) we get

$$Y_S = U_S\Gamma + V = \Phi\theta + V, \qquad (18)$$

the objective function can be written as

$$\begin{aligned} \operatorname{cov}(\hat{\theta}, S) &= \frac{1}{M}\sum_{i\in\hat{S}} r_i(\theta) r_i(\theta)^T = \frac{1}{M}\sum_{i=1}^{\mathcal{N}} r_i(\theta) G_i r_i(\theta)^T \\ &= \frac{1}{M}[Y - \Phi\theta]^T G [Y - \Phi\theta] \\ &= \frac{1}{M}[Y^T G Y - Y^T G \Phi\theta - \theta^T \Phi^T G Y + \theta^T \Phi^T G \Phi \theta] \end{aligned} \qquad (19)$$

where the $\mathcal{N} \times \mathcal{N}$ matrix G is define by $G = \operatorname{diag}\{G_1, \cdots, G_{\mathcal{N}}\}$ and $G_i = \{0, 1\}$. Using the weighted least squares estimator $\hat{\theta}_{WLS} = (\Phi^T G \Phi)^{-1}(\Phi^T G Y)$ equation (19) can be simplified as

$$\mathrm{cov}(\hat{\theta}, S) = \frac{1}{M}[Y^T G Y - Y^T G \Phi(\Phi^T G \Phi)^{-1}(\Phi^T G Y)] \quad (20)$$

it is easy to see that $\Phi^T G \Phi = \sum_{i=1}^{\mathcal{N}} \phi_i G_i \phi_i^T \to \sigma^2 I_p$, as $\mathcal{N} \to \infty$ so that the second term in (20) can be simplified as

$$\begin{aligned}
&\frac{1}{\sigma^2 M^2} Y^T G \Phi \Phi^T G Y = \frac{1}{\sigma^2 M^2} \sum_{i=1}^{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} (y_i G_i \phi_i^T)(\phi_k G_k y_k^T) \\
&= \frac{1}{\sigma^2 M^2} \sum_{i=1}^{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} \big((\theta^T \phi_i + v_i) G_i \phi_i^T\big)\big(\phi_k G_k (\theta^T \phi_k + v_k)^T\big) \\
&= \frac{1}{\sigma^2 M^2} \sum_{i=1}^{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} G_i \big(\theta^T \phi_i \phi_i^T \phi_k \phi_k^T \theta\big) G_k \\
&\quad + \frac{1}{\sigma^2 M^2} \sum_{i=1}^{\mathcal{N}} G_i (\theta^T \phi_i \phi_i^T) \sum_{k=1}^{\mathcal{N}} \phi_k v_k^T G_k \\
&\quad + \frac{1}{\sigma^2 M^2} \sum_{i=1}^{\mathcal{N}} G_i v_i \phi_i^T \sum_{k=1}^{\mathcal{N}} \phi_k \phi_k^T \theta G_k \\
&\quad + \frac{1}{\sigma^2 M^2} \sum_{i=1}^{\mathcal{N}} G_i v_i \phi_i^T \sum_{k=1}^{\mathcal{N}} \phi_k v_k^T G_k \qquad (21)
\end{aligned}$$

since v_i, ϕ_i and G_i are independent so that the last three terms vanishes in (21) when $N \to \infty$. It remains to analyze the first term

$$\begin{aligned}
&\frac{1}{\sigma^2 M^2} \sum_{i=1}^{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} G_i \big(\theta^T \phi_i \phi_i^T \phi_k \phi_k^T \theta\big) G_k \\
&\qquad = \frac{1}{\sigma^2 M^2} \sum_{i=1}^{\mathcal{N}} G_i \theta^T \phi_i \phi_i^T \sum_{k=1}^{\mathcal{N}} \phi_k \phi_k^T \theta G_k \qquad (22)
\end{aligned}$$

consider the first expression of left hand side of (22)

$$\frac{1}{\sigma^2 M} \theta^T \sum_{i=1}^{\mathcal{N}} G_i \phi_i \phi_i^T \to \frac{\sigma^2}{\sigma^2} \theta^T \qquad \text{if} \quad \mathcal{N} \to \infty$$

and similarly $\frac{1}{\sigma^2 M} \sum_{k=1}^{\mathcal{N}} \phi_k \phi_k^T \theta G_k$. So we conclude

$$\frac{1}{\sigma^2 M^2} Y^T G \Phi \Phi^T G Y \to \sigma^2 \|\theta\|^2$$

so that

$$\begin{aligned}
\det \mathrm{cov}(\hat{\theta}, S) &= \det[Y^T G Y - Y^T G \Phi \Phi^T G Y] \\
&= \det[\frac{1}{\sigma^2 \|\theta\|^2} Y^T G Y - I_p] \qquad (23)
\end{aligned}$$

let $M = \frac{1}{\sigma^2 \|\theta\|^2} Y^T G Y$, then $\min_{\theta,S} \det \mathrm{cov}(\hat{\theta}, S) = \min_{\theta,S} \det(M - I) = \min_{\theta,S} \prod_{k=1}^{n} (\lambda_k - 1)$, where λ_k is the kth eigenvalue of M. From the following series of implications,

$$\det \mathrm{cov}(\hat{\theta}, S) \geq 0, \to \lambda_k - 1 \geq 0, \quad k = 1, \cdots, n$$

to attain the minimum we have to minimize the eigenvalue of M. Now for any two positive matrices A_1 and A_2, such that $A_1 < A_2$ we have $\lambda_k(A_1) > \lambda_k(A_2)$, which proves the theorem.

Then, we have the following algorithm to compute the MCD estimate.

Random search algorithm 2

- **Step 1:** Arrange the squared outputs as $\|y\|_{[1]}^2 \leq \cdots \leq \|y\|_{[M]}^2 \leq \cdots \leq \|y\|_{[\mathcal{N}]}^2$.
- **Step 2:** Let $G^* = \mathrm{daig}\{G_1, \cdots, G_{\mathcal{N}}\}$ be defined as

$$G_i = \begin{cases} 1 & \text{if} \quad i = [1], [2], \cdots, [M], \\ 0 & \text{if} \quad i \neq [1], [2], \cdots, [M]. \end{cases}$$

- **Step 3:** Calculate the unknown parameters in the EIV model.

IV. CLOSED LOOP IDENTIFICATION WITH OUTLIERS

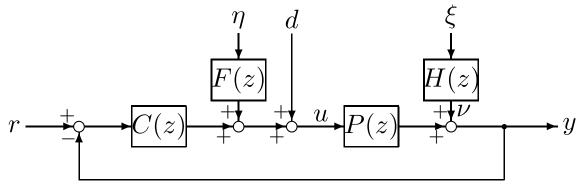

Fig. 3. Feedback system

Consider the problem of identifying the plant $P(z)$ and the controller $C(z)$ of the closed loop system shown in Fig. 3, where $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$ are the input-output signals of the plant. The unobserved disturbances $\eta \in \mathbb{R}^m$ and $\xi \in \mathbb{R}^p$ are white noise with zero mean and positive definite covariance matrices acting on the noise filters $H(z)$ and $F(z)$ respectively. The observed input signals $r \in \mathbb{R}^p$ and $d \in \mathbb{R}^m$ can be considered as a measurable disturbance. Let the plant and the control input be described by

$$y_t = P(z) u_t + H(z) \xi_t \qquad (24)$$

$$u_t = d_t + C(z)[r_t - y_t] + F(z) \eta_t \qquad (25)$$

where $P(z), H(z), C(z)$ and $F(z)$ are the transfer matrices of the plant, noise filter, controller and the measurement noise filter respectively.

Throughout this section, it is assumed that the closed loop system is internally stable, and the random processes d, r, η, ξ are wide-sense stationary, zero-mean and mutually uncorrelated.

Define the joint augmented input and output processes as

$$w := \begin{bmatrix} u \\ y \end{bmatrix} \in \mathbb{R}^l, \; v := \begin{bmatrix} d \\ r \end{bmatrix} \in \mathbb{R}^l, \text{ and } \mathcal{X} := \begin{bmatrix} \eta \\ \xi \end{bmatrix} \in \mathbb{R}^l. \qquad (26)$$

Then it can be shown that (24) and (25) can be written as [11],

$$x_{t+1} = A x_t + B_1 d_t + B_2 r_t + \begin{bmatrix} K_1 & K_2 \end{bmatrix} \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix}, \qquad (27)$$

$$\begin{bmatrix} y_t \\ u_t \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} x(t) + \begin{bmatrix} 0 & 0 \\ D_1 & D_2 \end{bmatrix} \begin{bmatrix} r_t^1 \\ r_t^2 \end{bmatrix} + \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix}. \qquad (28)$$

Since (27) and (28) are in the form of (5), we conclude that the results obtained for EIV problem can be applied for the closed loop system.

V. Numerical Example

Fig. 4 is sketch of the heat exchanger. As we can observe from the drawing the Hydro Carbon will pass inside the heat exchanger from tube inlet and leave from tube outlet and temperature recorded. This side of heat exchanger where the Hydro Carbon is passing called tube side.

On the other side of the heat exchanger where it called shell side. The steam is passing from the shell side between tubes to increase the temperature of the Hydro Carbon. The flow rate of the steam is unmeasured. Fig. 5, shows the input-output data of the heat exchanger. As it can be seen from Fig. 5, the input change the mean so that in order to detect the outliers we divide the input in to three parts. First part starts from $t = 1$ to $t = 940$, while the second from $t = 941$ to $t = 1162$ and part 3 from $t = 1163$ to $t = 2332$. Fig. 6, shows the residuals of the first and second attempt to detect the outliers in first part. Whereas Fig. 7 shows the residuals for the second and the third part. Finally Fig. 8, shows the Bode plot of the estimated heat exchanger system before and after detecting outliers.

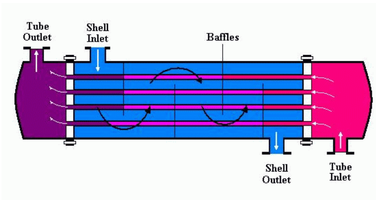

Fig. 4. Sketch of heat exchanger

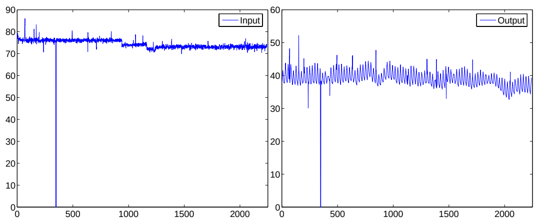

Fig. 5. Input-Output data for heat exchanger

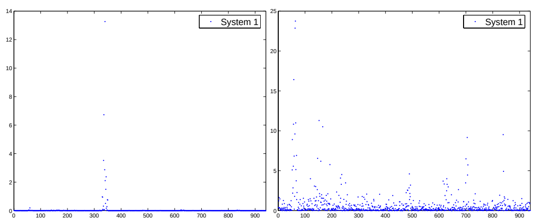

Fig. 6. detect outliers in part 1

VI. Conclusion

In this paper, we have considered the EIV identification problem with outliers in the framework of subspace method. We show that detecting the outliers in closed loop system is special case of the errors-in-variables problem. Moreover, we have propose the random search algorithm to solve the MCD problem. Then, we detect the outliers using the MCD method. The presented algorithm has been applied to heat exchanger data.

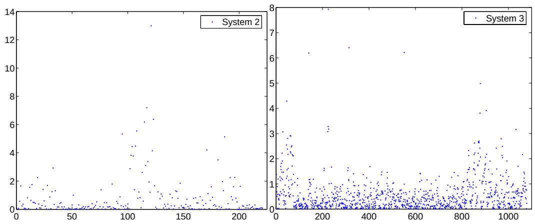

Fig. 7. detect outliers in part 2 and 3

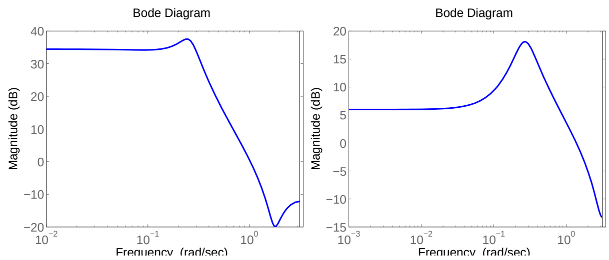

Fig. 8. Bode plots of the two-output system

References

[1] E. Bai: A random least-trimmed-squares identification algorithm; *Automatica*, Vol. 39, No. 9, pp. 1651-1659 (2003)

[2] G. Picci and T. Katayama: Stochastic realization with exogenous inputs and subspace methods identification; *Signal Processing*, Vol. 52, No. 2, pp. 145-160 (1996)

[3] C. Chou and M. Verhaegen: Subspace algorithms for the identification of multivariable dynamic errors-in-variables models; *Automatica*, Vol. 33, No. 10, pp. 1857-1869 (1997)

[4] T. Gustafsson: Subspace identification using instrumental variable techniques; *Automatica*, Vol. 37, No. 1, pp. 2005-2010 (2001)

[5] H. Tanaka and J. ALMutawa and T. Katayama: Subspace identification of linear systems with observation outliers; *Trans. Inst. Sys. Contr. and Inform. Eng.*, Vol. 17, No. 2, pp. 89-96 (2004)

[6] J. ALMutawa and H. Tanaka and T. Katayama: EM algorithm for system identification in the presence of outliers; *Trans. Inst. Sys. Contr. and Inform. Eng.*, Vol. 18, No. 5, pp. 146-151 (2005)

[7] P. J. Rousseeum, S. Van Aelst, K. Van Driessen and J. Aguíló: Robust multivariate regression; *Technometrics*, Vol. 46, No. 3, pp. 293–305 (2004)

[8] P. Rousseeum: Least median of squares regression; *J. American Statistical Assoc.*, Vol. 79, pp. 871-880 (1984)

[9] D. L. Donoho and P. J. Huber: The notion of breakdown point. In P. Bickel, K. Doksum and J.L. Hodges (Eds.), *A Festschrift for Erich Lehmann. Wadsworth, CA*, pp. 157–184 (1983)

[10] J. Wang and S.J. Qin: A new subspace identification approach based on principal component analysis; *J. Process Control*, Vol. 12, pp. 841–855 (2002)

[11] T. Katayama, H. Kawauchi and G. Picci: Subspace identification of closed loop systems by the orthogonal decomposition method; *Automatica*, Vol. 41, No. 2, pp. 863-872 (2005)